Weekly Report

1 Done

1.1 Vast Presentation

I modified the presentation slides according to the advices given in the group meeting.

1.2 Article about Visual privacy preservation

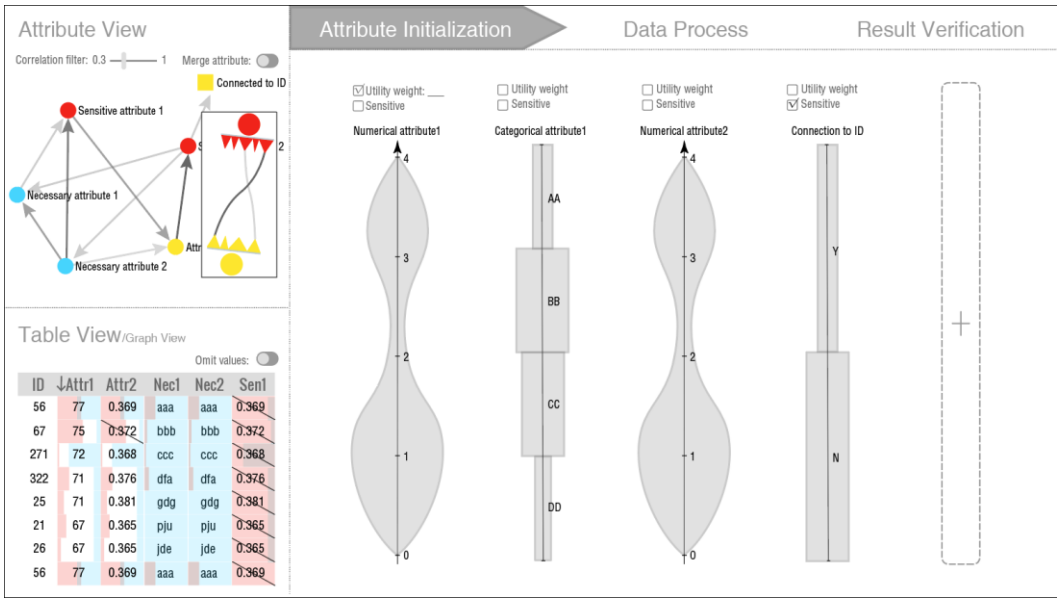
I listed the outline and started writing.

1.3 Idea

The following three pictures are my designs for our interface.

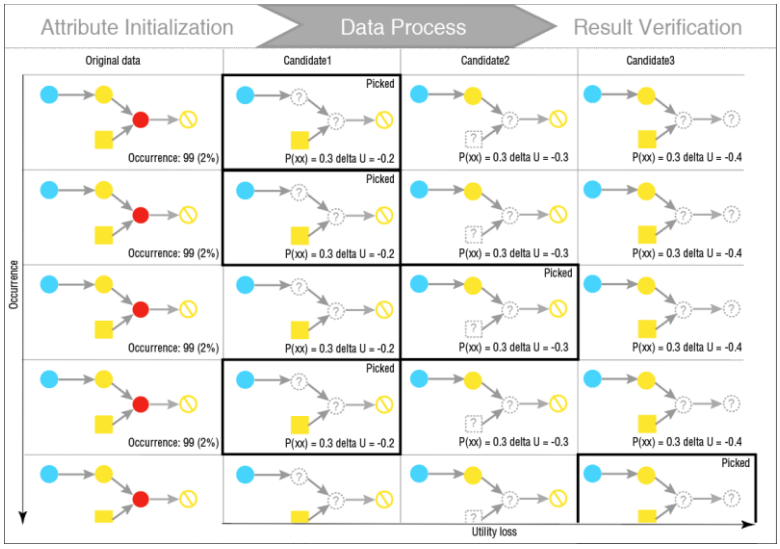
The two views on the left are fixed. The attribute view shows the correlation between attributes by node-link graph. The bottom-left view is for datasets. Users can check the datasets in the forms of table of graph.

The view on the right is the main view. It will change according to the analysis stages. We have three stages totally.



Stage 1: Attribute initialization.

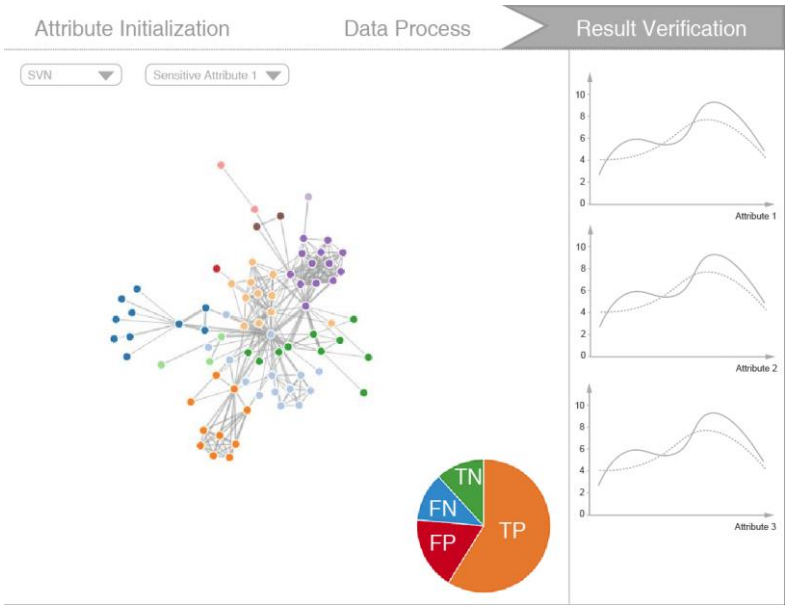
- First, users define sensitive attribute and set utility weight for attributes.
- Then, for each attribute, users specify events by attribute values. Here, users can create attributes by the connections between nodes and special nodes, like high degree. The events could be the individual is less than 18-year-old, or she/he follows Donald Trump, or she/he follows three pop stars.
- According to the entire datasets, we can generate a General Bayesian Network. It will be visualized in the attribute view.



Stage 2: Data process

Even if we remove sensitive information from the data, adversaries still have chances to infer sensitive information through other attributes. So, our target is to reduce their accuracy. For instance, event A has a probability of 0.4 in the entire datasets. Then, for any individual T, adversaries should infer that the probability of event A happening to T is $0.4 \pm \varepsilon$, here ε is a user-defined parameter. To preserve privacy, we remove information. The events with low utility weight and very relevant to sensitive information will be prioritized. Moreover, to preserve the data distribution, the utility weight will be improved if a large number of same events are removed.

- Here, we cluster data records by the events they have. The clusters will be shown as partial General Bayesian Network on the left. The clusters are sort by occurrences. On the right, system provides three options for privacy preservation. The candidates are sort by utility loss.
- System will choose the first option with the lowest utility loss by default. Users can change the option by clicking other option.



Stage 3: Result verification

- Users can verify the privacy preservation by employing other attack models like SVM or TAN. The results will be shown by pie chart for summary, node-link graph and table for details. Moreover, the modification of distribution will be shown as line charts on the right.

  We will start coding next week. Yanling and me will focus on front end. Wenlong and Rusheng will focus on back end.

2 Work Hours

Monday	9:00-11:30	12:30-5:00	6:00-11:40
Tuesday	9:00-11:30	12:30-5:00	6:00-8:30
Wednesday	9:00-11:30	12:30-5:00	6:00-8:30
Thursday	9:00-11:30	12:30-5:00	6:00-8:30
Friday	9:00-11:30	12:30-5:00	6:00-11:00
Saturday	9:30-11:30	12:30-5:00	6:00-8:30
Sunday	10:00-11:00	12:00-5:40	6:20-8:30

3 Progress

Item	Deadline	Current progress	Remark
Vis presentation	10.24	Script is done.	
Go abroad	11.18	Ready for buying the ticket.	
Privacy program	10.31	Ready to code.	
Article	10.30	Start writing	